THE DISTANCE APPROACH TO APPROXIMATE COMBINATORIAL COUNTING

Alexander Barvinok and Alex Samorodnitsky

May 2000

Abstract. We develop general methods to obtain fast (polynomial time) estimates of the cardinality of a combinatorially defined set via solving some randomly generated optimization problems on the set. Geometrically, we estimate the cardinality of a subset of the Boolean cube via the average distance from a point in the cube to the subset. As an application, we present a new randomized polynomial time algorithm which approximates the permanent of a 0-1 matrix by solving a small number of Assignment problems.

1. Introduction

A general problem of combinatorial counting can be stated as follows: given a family $\mathcal{F} \subset 2^X$ of subsets of the ground set X, compute or estimate the cardinality $|\mathcal{F}|$ of the family. We would like to do the computation efficiently, *in polynomial time*. Of course, one should clarify what "given" means, especially since in most interesting cases $|\mathcal{F}|$ is exponentially large in the cardinality $|X|$ of the ground set. Following the earlier paper [Barvinok 97a], we assume that the family $\mathcal{F}$ is defined by its *Optimization Oracle*:

(1.1) Optimization Oracle defining a family $\mathcal{F} \subset 2^X$

Input: A set of integer weights $\gamma_x : x \in X$.

Output: The number $\displaystyle\min_{Y \in \mathcal{F}} \sum_{x \in Y} \gamma_x$.

That is, for any given integer weighting $\{\gamma_x\}$ on the set X, we should be able to produce the minimum weight of a subset $Y \in \mathcal{F}$. As is discussed in [Barvinok 97a], for many important families $\mathcal{F}$ the Optimization Oracle is readily available. The following example is central for this paper.

Key words and phrases. combinatorial counting, permanent, Hamming distance, polynomial time algorithms, isoperimetric inequalities, Boolean cube.

The research of the first author was partially supported by NSF Grant DMS 9734138.

(1.2) Example: Perfect matchings in a graph. Let $G = (V, E)$ be a graph with the set V of vertices and set E of edges. We assume that G has no loops (edges whose endpoints coincide) and no isolated vertices. A set $M \subset E$ of edges is called a *matching* in G if every vertex of G is incident to at most one edge from M. A matching M is called *perfect* if every vertex of G is incident to precisely one edge from M. Let $\mathcal{F} \subset 2^E$ be the set of all *perfect matchings* in G. The problem of computing or estimating $|\mathcal{F}|$ efficiently is one of the hardest and most intriguing problems of combinatorial counting, see, for example, [Lovász and Plummer 86], [Jerrum and Sinclair 89], [Jerrum 95] and [Jerrum and Sinclair 97].

We observe that Optimization Oracle 1.1 can be efficiently constructed. Indeed, if we assign integer weights γ_e: $e \in E$ to the edges of the graph, the minimum weight of a perfect matching can be computed in $O(|V|^3)$ time, see, for example Section 11.3 of [Papadimitriou and Steiglitz 98].

A particularly interesting case is that of a *bipartite* graph G when the vertices of G are partitioned into two classes, $V = V^+ \cup V^-$ such that every edge $e \in E$ has one endpoint in V^+ and the other in V^-. Then the number of perfect matchings in G is equal to the permanent of a 0-1 matrix associated with G (see also Section 5). In this case, the corresponding optimization problem is known as the Assignment Problem. It is not only "theoretically easy", but in practice large instances are routinely solved as the Assignment Problem is a particular case of the minimum cost network flow problem (see, for example, Section 11.2 of [Papadimitriou and Steiglitz 98]).

Other interesting and generally difficult problems of combinatorial counting where the Optimization Oracle is provided by classical combinatorial optimization algorithms include counting bases in matroids, counting independent sets in matroids and counting bases in the intersection of two matroids over the same ground set, see [Jerrum and Sinclair 97] for a discussion of the counting problems and [Papadimitriou and Steiglitz 98] for a description of the underlying optimization algorithms. Particularly interesting special cases of those problems include counting spanning trees, counting forests and counting spanning subgraphs in a given graph and counting non-degenerate maximal minors in a given rectangular matrix over $GF(2)$. Some of the problems, such as counting spanning trees, admit a simple and efficient solution, others, such as counting matchings of all sizes in a graph, are known to be hard to solve exactly but can be solved approximately and still others, such as counting bases in matroids, are solved only in special cases. The problem of counting perfect matchings in a given graph, arguably the most famous problem of them all, still resists all attempts to solve it in full generality (see also Section 5).

The most general approach to combinatorial counting has been via Monte Carlo method. The key component of the method is the ability to sample a random point from the (almost) uniform distribution on $\mathcal{F}$. Often, to achieve this, a Markov chain on the set $\mathcal{F}$ is generated, so that it converges rapidly to the uniform distribution on $\mathcal{F}$ (see [Jerrum and Sinclair 97] for a survey). This approach resulted, for example,

in finding a polynomial time randomized algorithm to count matchings of all sizes in a given graph with a prescribed accuracy [Jerrum and Sinclair 89]. When the Markov chain approach works, it produces incomparably better results than the method of this paper. However, for many important counting problems, some of which are mentioned above, it is either not clear how to generate a rapidly mixing Markov chain or, when there is a "natural" candidate, it seems to be extremely hard to prove that the chain is indeed converging rapidly enough to the steady state (cf. [Jerrum and Sinclair 97]). In contrast, our approach produces very crude bounds, but it is totally insensitive to the fine structure of $\mathcal{F}$, so it is ready to handle a broad class of problems. In [Barvinok 97a], it was shown that the method allows one to decide whether the size $|\mathcal{F}|$ is exponentially large in the size $|X|$ of the ground set in some precisely defined sense. In this paper, we improve the estimates of [Barvinok 97a] in several directions and apply them to new problems, notably to the problem of estimating the permanent of a given 0-1 matrix.

The main idea of our approach is as follows. Given a family $\mathcal{F}$, we identify it with a subset F of a metric space (Ω, d), such that for any given point $x \in \Omega$ the distance $d(x, F) = \min\limits_{y \in F} d(x, y)$ can be quickly computed using Optimization Oracle 1.1 for $\mathcal{F}$. Then we estimate the cardinality $|F|$ from the distance $d(x, F)$ for a typical $x \in \Omega$. Intuitively, if $|F|$ is small, we expect the distance $d(x, F)$ from a random point $x \in \Omega$ to be large and vice versa. In this paper, Ω is the Boolean cube $\{0, 1\}^n$ and d is either the Hamming distance or its modification, although as we discuss in Section 7, some other possibilities may be of interest. Thus our approach can be considered as a refinement of the classical Monte-Carlo method: we do not only register how often a randomly sampled point $x \in \Omega$ lands in the target set F, but also take into account the distance $d(x, F)$. This allows us to get non-trivial bounds even when $|F|$ is exponentially small with respect to $|\Omega|$ so that x typically misses F.

The paper is organized as follows.

In Section 2, we introduce a "geometric cousin" of Optimization Oracle 1.1. Distance Oracle 2.2 describes a subset F of the Boolean cube $\{0, 1\}^n$ by computing a suitably defined distance d from a given point in the cube to the set. We show how to construct embeddings $\phi : \mathcal{F} \longrightarrow \{0, 1\}^n$, so that the Distance Oracle for the image $F = \phi(\mathcal{F})$ is derived from the Optimization Oracle for $\mathcal{F}$. We show that in some important cases (for example, when $\mathcal{F}$ is the set of perfect matchings in a graph), we can "squeeze" $\mathcal{F}$ into a substantially smaller cube than we would have expected for a general family $\mathcal{F}$.

In Section 3, we describe the bounds obtained by choosing d to be the Hamming distance in the cube. The bounds are sharp, meaning that we can't possibly estimate (in polynomial time) the cardinality of a subset $F \subset \{0, 1\}^n$ better if the only information available is the Hamming distance from any given point $a \in \{0, 1\}^n$ to the set F. Remarkably, the lower and the upper bound for $\alpha = n^{-1} \log_2 |F|$ converge when $\alpha \approx 0$ or $\alpha \approx 1$ and diverge the greatest when $\alpha = 1/2$.

In Section 4, we describe how to get better bounds for small sets by using a

suitably defined "randomized Hamming distance", which ignores a (random) part of the information contained in the standard Hamming distance. The isoperimetric problems arising here seem to be interesting in their own right. The proofs are not complicated but somewhat lengthy and therefore postponed till Section 6.

In Section 5, we apply our methods to get a new polynomial time algorithm to approximate the permanent of a given 0-1 matrix. Geometrically, we represent the set of the perfect matchings of the underlying bipartite graph on $n+n$ vertices as a subset F of the Boolean cube $\{0,1\}^m$ with $m = O(n \ln n)$ and estimate $|F|$ from the Hamming distance of a random point in the cube to F. We find the distance in question by averaging solutions of some randomly generated Assignment problems. We compare our method with other algorithms available in the literature. In particular, we show that our method allows us to recognize $n \times n$ matrices whose permanents are subexponential in n (Corollary 5.4).

In Section 6, we supply proofs of the results of Section 4.

In Section 7, we discuss possible ramifications of our approach and its relations with the Monte-Carlo method.

2. Distance Oracle and Cubical Embeddings

The idea of our method is to represent $\mathcal{F}$ geometrically as a subset F of the Boolean cube and then derive estimates of $|\mathcal{F}|$ using the average distance from a point in the cube to F.

(2.1) Definitions. Let $C_n = \{0,1\}^n$ be the Boolean cube and let dist be the Hamming distance in C_n, that is

$$\operatorname{dist}(a,b) = \sum_{i: \alpha_i \neq \beta_i} 1 \qquad \text{for} \quad a = (\alpha_1, \ldots, \alpha_n), \ b = (\beta_1, \ldots, \beta_n) \in C_n.$$

More generally, let us fix n functions $d_i : \{0,1\} \times \{0,1\} \longrightarrow \mathbb{Z}$, $i = 1, \ldots, n$, which we interpret as *penalties*. We assume that $d_i \geq 0$ and that $d(0,0) = d(1,1) = 0$. Finally, let

$$d(a,b) = \sum_{i=1}^{n} d_i(\alpha_i, \beta_i), \quad \text{where} \quad a = (\alpha_1, \ldots, \alpha_n) \quad \text{and} \quad b = (\beta_1, \ldots, \beta_n)$$

be the L^1 distance function determined by the penalties $\{d_i\}$.

If $d_i(\alpha, \beta) = 1$ whenever $\alpha \neq \beta$ then $d(a,b) = \operatorname{dist}(a,b)$.

For a subset $B \subset C_n$ and a point $a \in C_n$, let

$$d(a,B) = \min_{b \in B} d(a,b)$$

be the distance from a to B. In particular, let

$$\operatorname{dist}(a,B) = \min_{b \in B} \operatorname{dist}(a,b)$$

be the Hamming distance from a point a to the subset B.

We will be working with the following "geometric cousin" of Optimization Oracle 1.1.

(2.2) Distance Oracle defining a set $F \subset C_n$

Input: A point $a \in C_n$ and penalties $d_i : \{0,1\} \times \{0,1\} \longrightarrow \mathbb{Z}$, $i = 1, \ldots, n$.

Output: The number $d(a, F)$.

There is an obvious way to associate with a family $\mathcal{F} \subset 2^X$ a subset $F \subset C_{|X|}$ of the Boolean cube.

(2.3) Straightforward embedding. Let us identify the ground set X with the set $\{1, \ldots, n\}$, $n = |X|$. Let $\mathcal{F}$ be a family of subsets of $\{1, \ldots, n\}$ given by its Optimization Oracle. For a subset $Y \in \mathcal{F}$ let us define the indicator $y \in C_n$, $y = (\eta_1, \ldots, \eta_n)$ by

$$\eta_i = \begin{cases} 1 & \text{if } i \in Y \\ 0 & \text{if } \ i \notin Y. \end{cases}$$

Let $F = \big\{ y \in C_n : Y \in \mathcal{F} \big\}$ be the set of all indicators of subsets $Y \in \mathcal{F}$.

Let us construct the Distance Oracle for the set $F \subset C_n$. Given a point $a = (\alpha_1, \ldots, \alpha_n) \in C_n$ and penalties d_i, $i = 1, \ldots, n$, let us define weights γ_i by $\gamma_i = d_i(\alpha_i, 1) - d_i(\alpha_i, 0)$. Then for a set $Y \subset \{1, \ldots, n\}$ and its indicator $y = (\eta_1, \ldots, \eta_n) \in C_n$, we have

$$\sum_{i \in Y} \gamma_i = \sum_{i \in Y} \big(d_i(\alpha_i, 1) - d_i(\alpha_i, 0) \big) = \sum_{i=1}^n d_i(\alpha_i, \eta_i) - \sum_{i=1}^n d_i(\alpha_i, 0) = d(a, y) - d(a, 0).$$

Hence, given the output

$$\lambda = \min_{Y \in \mathcal{F}} \sum_{i \in Y} \gamma_i$$

of Oracle 1.1 for the family $\mathcal{F}$, we can easily compute the output

$$d(a, F) = \lambda + d(a, 0)$$

of Oracle 2.2 for the set F. Thus, given an Optimization Oracle 1.1 for a family $\mathcal{F} \subset 2^X$, we can efficiently construct a Distance Oracle 2.2 for a set $F \subset C_n$, $n = |X|$, such that $|F| = |\mathcal{F}|$.

To be able to estimate the cardinality $|\mathcal{F}|$ with a better precision, we would like to embed $\mathcal{F}$ into a smaller Boolean cube. Sometimes this is indeed possible.

(2.4) Economical embedding. Suppose that the ground set X can be represented as a union $X = X_1 \cup \ldots \cup X_k$ of (not necessarily disjoint) parts X_i, so that $|Y \cap X_i| = 1$ for every subset $Y \in \mathcal{F}$ and every X_i. In other words, every member of $\mathcal{F}$ is a transversal of the cover of X by $X_1, \ldots, X_k$. Let

$$m_i = \lceil \log_2 |X_i| \rceil \quad \text{and} \quad m = \sum_{i=1}^k m_i.$$

We construct an embedding $\mathcal{F} \longrightarrow C_m$ as follows.

First, we index the elements of X_i by distinct binary strings of length m_i, that is, we choose an embedding $\phi_i : X_i \longrightarrow C_{m_i}$. Thus for any $x \in X_i$ the point $\phi_i(x)$ is a binary string of length m_i and $\phi_i(x) \neq \phi_i(y)$ provided $x \neq y$.

Let us identify

$$C_m = C_{m_1} \times \ldots \times C_{m_k}.$$

For a subset $Y \in \mathcal{F}$, let us define $y \in C_m$ as

$$y = (y_1, \ldots, y_k), \quad \text{where} \quad y_i = \phi_i(Y \cap X_i) \in C_{m_i}.$$

Note that y is well-defined, since every intersection $Y \cap X_i$ consists of a single point. Let $F = \big\{ y \in C_m : Y \in \mathcal{F} \big\}$. Clearly, $|F| = |\mathcal{F}|$.

Given an Optimization Oracle 1.1 for $\mathcal{F}$, let us construct a Distance Oracle 2.2 for F. The input of Oracle 2.2 consists of a point $a \in C_m$ (binary string of length m) and penalty functions $\{d_i : i = 1, \ldots m\}$. We view a as

$$a = (a_1, \ldots, a_k), \quad \text{where} \quad a_i \in C_{m_i}.$$

The penalties d_i, $i = 1, \ldots, m$ give rise to the L^1 distance function d on binary strings, cf. Definition 2.1. For a point $x \in X$, let us define its weight γ_x by

$$\gamma_x = \sum_{i:\ x \in X_i} d\big(a_i, \phi_i(x)\big). \tag{2.4.1}$$

Let $Y \in \mathcal{F}$ be a set and let $y \in C_m$ be the point representing Y. We observe that

$$\sum_{x \in Y} \gamma_x = \sum_{x \in Y} \sum_{i:\ x \in X_i} d\big(a_i, \phi_i(x)\big) = \sum_{i=1}^{k} d(a_i, y_i) = d(a, y).$$

Hence, the outputs of Oracles 1.1 and 2.2 coincide:

$$\min_{Y \in \mathcal{F}} \sum_{x \in Y} \gamma_x = \min_{y \in Y} d(a, y).$$

Thus, given an Optimization Oracle 1.1 for a family $\mathcal{F} \subset 2^X$, we can efficiently construct a Distance Oracle 2.2 for a set $F \subset C_m$, such that $|F| = |\mathcal{F}|$. More precisely, given a point $a \in C_m$ and penalties $\{d_i\}$, we compute weights $\{\gamma_x\}$ on X by (2.4.1) in $O(k|X| \ln |X|)$ time and then apply Optimization Oracle 1.1 to find the minimum weight λ of a subset $Y \in \mathcal{F}$ in this weighting. The distance $d(a, F)$ is equal to λ.

(2.5) Example: Embedding perfect matchings. Let $\mathcal{F}$ be the family of all perfect matchings in a graph $G = (V, E)$, see Example 1.2. The straightforward embedding (2.3) identifies $\mathcal{F}$ with a subset F of the Boolean cube $\{0,1\}^{|E|}$ and provides us with Distance Oracle 2.2 for F. We will be better off using the economical embedding (2.4). Indeed, for a vertex $v \in V$ of G, let E_v be the set of edges of G incident to v. Then $E = \cup_{v \in V} E_v$ and every perfect matching has exactly one edge in every set E_v. Hence the embedding (2.4) identifies $\mathcal{F}$ with a subset F of the Boolean cube $\{0,1\}^m$, where

$$m = \sum_{v \in V} \lceil \log_2 |E_v| \rceil$$

and provides us with Distance Oracle 2.2 for F. Given a point $a \in C_m$, by (2.4.1) we compute weights γ_e on the edges E in $O(|E| \ln |E|)$ time (since every edge $e \in E$ belongs to exactly two sets E_v) and then find the minimum weight λ of a perfect matching in G in $O(|V|^3)$ time. The distance $d(a, F)$ from a to F is equal to λ.

Typically, if the graph has $|V| = n$ vertices and $\Omega(n^2)$ edges, the dimension of the straightforward embedding will be $\Omega(n^2)$, whereas the dimension of the economical embedding will be $O(n \ln n)$. We observe that for bipartite graphs we can reduce the dimension further by a factor of 2 at least by choosing

$$m = \min\Big\{ \sum_{v \in V^+} \lceil \log_2 |E_v| \rceil, \sum_{v \in V^-} \lceil \log_2 |E_v| \rceil \Big\},$$

since every perfect matching $M \subset E$ will be a transversal of either partition $E = \cup_{v \in V^+} E_v$ or $E = \cup_{v \in V^-} E_v$.

Another natural case of economical embedding 2.4 arises when $\mathcal{F}$ is the set of common bases of two matroids on the same ground set, one of which is a transversal matroid. It would be interesting to find out if similar economical embeddings can be constructed for a broader class of families $\mathcal{F} \subset 2^X$, for example, when $\mathcal{F}$ consists of "small" sets, that is, when $|Y| << |X|$ for any $Y \in \mathcal{F}$.

3. Estimating Cardinality from the Hamming Distance

In this section, we obtain estimates of the cardinality of a subset $F \subset C_n$ if we choose $d_i(0,1) = d_i(1,0) = 1$, $i = 1, \ldots, n$ in Distance Oracle 2.2. In other words, we estimate $|F|$, provided we can compute the Hamming distance $\operatorname{dist}(x, F)$ to F from any given point $x \in C_n$, cf. Definitions 2.1. Our main tool is the *average* Hamming distance from a point to the set.

(3.1) Definition. Let $A \subset C_n$ be a subset of the Boolean cube. Let

$$\Delta(A) = \frac{1}{2^n} \sum_{x \in C_n} \operatorname{dist}(x, A)$$

be the average Hamming distance from a point in the cube to the set A.

Obviously, $\Delta(A) \leq \Delta(B)$ if $B \subset A$.

(3.2) Example: Set consisting of a single point. Suppose that the set A is a point. Without loss of generality we assume that $A = \{(0, \ldots, 0)\}$. Then, for $x = (\xi_1, \ldots, \xi_n)$ we have $\text{dist}(x, A) = \text{dist}(x, 0) = \xi_1 + \ldots + \xi_n$ and

$$\Delta(A) = \frac{1}{2^n} \sum_{x \in C_n} \text{dist}(x, A) = \frac{1}{2^n} \sum_{x \in C_n} (\xi_1 + \ldots + \xi_n) = \frac{n}{2}.$$

It follows then that $\Delta(A) \leq n/2$ for any non-empty $A \subset C_n$ and that $\Delta(A) = n/2$ if and only if A consists of a single point.

Our first objective is to present a probabilistic algorithm that computes $\Delta(A)$ approximately by averaging $\text{dist}(x, A)$ for a number of randomly chosen $x \in C_n$.

(3.3) Algorithm for computing $\Delta(A)$

Input: A set $A \subset C_n$ defined by its Distance Oracle 2.2 and a number $\epsilon > 0$.

Output: A number α approximating $\Delta(A)$ within error ϵ.

Algorithm: Let $k = \lceil 48n/\epsilon^2 \rceil$. Sample k points $x_1, \ldots, x_k \in C_n$ independently at random from the uniform distribution in the cube C_n. Apply Distance Oracle 2.2 to find $\text{dist}(x_i, A)$, $i = 1, \ldots, k$. Compute $\alpha = \dfrac{1}{k} \displaystyle\sum_{i=1}^{k} \text{dist}(x_i, A)$. Output α.

To prove that Algorithm 3.3 indeed approximates $\Delta(A)$ with the desired accuracy, we need a couple of technical results. The first lemma supplies us with important *concentration inequalities* for the Boolean cube.

(3.4) Lemma. *Let $C_N = \{0, 1\}^N$ be the Boolean cube and let $f : C_N \longrightarrow \mathbb{R}$ be a function such that*

$$|f(x) - f(y)| \leq \text{dist}(x, y) \quad \textit{for all} \quad x, y \in C_N.$$

Let

$$\mathbf{E}\,(f) = \frac{1}{2^N} \sum_{x \in C_N} f(x)$$

be the average value of f. Let $\mathbf{P}$ denote the uniform probability measure on C_N, so $\mathbf{P}\,(A) = |A|/2^N$ for a set $A \subset C_N$.

Then for any $\delta > 0$

$$\mathbf{P}\,\Big\{x \in C_N : |f(x) - \mathbf{E}\,(f)| \geq \delta\Big\} \leq 2 \exp\Big\{\frac{-\delta^2}{16N}\Big\}.$$

Proof. See Sections 6.2 and 7.9 of [Milman and Schechtman 86]. □

The next lemma provides a useful "scaling" trick.

(3.5) Lemma. *Let us fix positive integers k and n and let $N = kn$. Let us identify $C_N = C_n \times \ldots \times C_n = (C_n)^k$. Thus a point $x \in C_N$ is identified with a k-tuple $x = (x_1, \ldots, x_k)$, where $x_i \in C_n$ for $i = 1, \ldots, k$.*

For a subset $A \subset C_n$, let $B = A \times \ldots \times A = A^k \subset C_N$. Then

$$\operatorname{dist}(x, B) = \sum_{i=1}^{k} \operatorname{dist}(x_i, A) \quad \textit{for any} \quad x = (x_1, \ldots, x_k) \in C_N$$

and

$$\Delta(B) = k\Delta(A).$$

Proof. Clearly,

$$\operatorname{dist}(x, y) = \sum_{i=1}^{k} \operatorname{dist}(x_i, y_i) \quad \text{for all} \quad x, y \in C_N,$$

hence the first identity follows. Next,

$$\begin{aligned}\Delta(B) = &\frac{1}{2^N} \sum_{x \in C_N} \operatorname{dist}(x, B) = \frac{1}{2^N} \sum_{x_1, \ldots, x_k \in C_n} \sum_{i=1}^{k} \operatorname{dist}(x_i, A) \\ = &\frac{k 2^{n(k-1)}}{2^{nk}} \sum_{x \in C_n} \operatorname{dist}(x, A) = \frac{k}{2^n} \sum_{x \in C_n} \operatorname{dist}(x, A) = k\Delta(A).\end{aligned}$$

□

Now we can prove correctness of Algorithm 3.3.

(3.6) Theorem. *With probability at least 0.9, the output α of Algorithm 3.3 satisfies the inequality $|\Delta(A) - \alpha| \leq \epsilon$.*

Proof. Let $N = nk$ and let us identify $C_N = (C_n)^k$ as in Lemma 3.5. Let $B = A^k \subset C_N$. Let $f : C_N \longrightarrow \mathbb{R}$ be defined by $f(x) = \operatorname{dist}(x, B)$. Applying Lemma 3.4 with $\delta = k\epsilon$ and observing that $\mathbf{E}\,(f) = \Delta(B)$, we conclude that

$$\mathbf{P}\,\Big\{x : |\operatorname{dist}(x, B) - \Delta(B)| \geq k\epsilon\Big\} \leq 2\exp\Big\{-\frac{(\epsilon k)^2}{16N}\Big\} = 2\exp\Big\{-\frac{\epsilon^2 k}{16n}\Big\} \leq 0.1.$$

Since by Lemma 3.5

$$\Delta(B) = k\Delta(A) \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{k} \operatorname{dist}(x_i, A) = \frac{1}{k}\operatorname{dist}(x, B)$$

for $x = (x_1, \ldots, x_k)$, we conclude that

$$\begin{gathered}\mathbf{P}\left\{x_1, \ldots, x_k : \Big|\frac{1}{k}\sum_{i=1}^{k} \operatorname{dist}(x_i, A) - \Delta(A)\Big| \geq \epsilon\right\} = \\ \mathbf{P}\Big\{x : |\operatorname{dist}(x, B) - \Delta(B)| \geq k\epsilon\Big\} \leq 0.1,\end{gathered}$$

and the proof follows. $\square$

Remark. Hence to evaluate $\Delta(A)$ within error ϵ we have to average $O(n\epsilon^{-2})$ values $\operatorname{dist}(x_i, A)$. By doing that, we allow probability 0.1 of failure. As usual, to attain a lower probability $\delta > 0$ of failure, one should run Algorithm 3.3 $O(\ln \delta^{-1})$ times and then select the median of the computed α's (cf. [Jerrum *et al.* 86]). For all applications, choosing $\epsilon = 1$ will suffice and in many cases $\epsilon = \sqrt{n}$ will do (cf. Section 5 and [Barvinok 97a]). Hence, often we will have to apply Oracle 2.2 only a constant number of times.

We would like to relate the value of $\Delta(A)$ to the cardinality $|A|$.

(3.7) Definition. Entropy Function. For $0 \leq x \leq 1/2$ let

$$H(x) = x \log_2 \frac{1}{x} + (1-x) \log_2 \frac{1}{1-x}.$$

We agree that $H(0) = 0$. Thus H is an increasing concave function on the interval $[0, 1/2]$.

We use the following estimate (see, for example, Theorem 1.4.5 of [van Lint 99])

$$\sum_{k=0}^{r} \binom{n}{k} \leq 2^{nH(r/n)} \qquad \text{for} \quad r \leq n/2. \tag{3.7.1}$$

Also, we remark that around $x = +0$ we have

$$H(x) = x \log_2 \frac{1}{x} + O(x) \quad \text{and} \quad H\left(\frac{1}{2} - x\right) = 1 - \frac{2}{\ln 2}x^2 + O(x^3) \tag{3.7.2}$$

We will use the classical isoperimetric inequality for the Boolean cube (see, for example, [Leader 91]).

(3.8) Harper's Theorem. *Let $A \subset C_n$ be a set such that*

$$|A| \geq \sum_{k=0}^{r} \binom{n}{k}$$

for some integer r. Then, for any non-negative integer t

$$\Big|\Big\{x \in C_n : \operatorname{dist}(x, A) \leq t\Big\}\Big| \geq \sum_{k=0}^{r+t} \binom{n}{k}.$$

We are going to obtain an estimate of the cardinality of a set $A \subset C_n$ in terms of the average Hamming distance $\Delta(A)$ from a point $x \in C_n$ to A. It is convenient to express the estimate in terms of a related quantity

$$\rho = \rho(A) = \frac{1}{2} - \frac{\Delta(A)}{n}.$$

As follows from Example 3.2, for every non-empty set $A \subset C_n$ we have $0 \leq \rho(A) \leq 1/2$. We observe that $\rho(A) = 0$ if and only if A consists of a single point and that $\rho(A) = 1/2$ if and only if A is the whole cube C_n.

(3.9) Theorem. *Let $A \subset C_n$ be a non-empty set. Let*

$$\rho = \frac{1}{2} - \frac{\Delta(A)}{n}.$$

Then

$$1 - H\Big(\frac{1}{2} - \rho\Big) \leq \frac{\log_2 |A|}{n} \leq H(\rho).$$

Before we proceed with a formal proof, we would like to highlight some ideas.

(3.10) The idea of the proof. Extremal sets. Let $A \subset C_n$ be a set. Concentration inequalities (Lemma 3.4) imply that the average distance $\Delta(A)$ is approximately equal to the distance $\operatorname{dist}(x, A)$ from a "typical" point $x \in C_n$ to A. For a given positive integer t, let us consider the t-neighborhood $A_t = \big\{x \in C_n : \operatorname{dist}(x, A) \leq t\big\}$ of A. We expect that $\Delta(A) \approx t_1$, where t_1 is the smallest value of t such that A_t covers "almost all" cube C_n. The neighborhood A_t grows the slowest when A is a ball in the Hamming metric, that is when $A = \big\{x : \operatorname{dist}(x, x_0) \leq r\big\}$ for some $x_0 \in C_n$ and some $r > 0$, as follows from Harper's Theorem 3.8, cf. also [Leader 91]. Hence the upper bound for $n^{-1} \log_2 |A|$ in Theorem 3.9 is attained (up to an $O(n^{-1/2})$ error term) when A is a ball. The neighborhood A_t grows the fastest when the points of A are spread around in C_n. In any case, the size $|A_t|$ does not exceed the sum of sizes of the balls of radius t centered at the points of A. Thus the lower bound for $n^{-1} \log_2 |A|$ in Theorem 3.9 is obtained from this "packing" type argument. One can show that if the points of A are chosen at random in C_n, then with high probability the lower bound is indeed attained asymptotically. More precisely, let us fix a number $0 < \beta < 1$ and let A be the set of $\lfloor 2^{\beta n} \rfloor$ points chosen at random from C_n. Then with the probability that tends to 1 as n grows to infinity, $\beta = 1 - H\big(\frac{1}{2} - \rho\big) + O(n^{-1/2})$. The proof is straightforward, but technical and therefore omitted.

Finally, we note that using average distance $\Delta(A)$ and the scaling trick (Lemma 3.5) allows us to get rid of $O(n^{-1/2})$ error terms in the proof.

Proof of Theorem 3.9. Let us choose a positive even integer m, let $N = mn$ and let us identify $C_N = (C_n)^m$, as in Lemma 3.5. Let $B = A^m \subset C_N$. Let us fix the uniform probability measure $\mathbf{P}$ on C_N.

Let $\alpha = \log_2 |A|/n$, so $|A| = 2^{\alpha n}$ and $|B| = 2^{\alpha N}$. Let $0 \leq \gamma \leq 1/2$ be a number such that $H(\gamma) = \alpha$ and let $r = \lfloor N\gamma \rfloor$. Then by (3.7.1)

$$|B| = 2^{N \cdot H(\gamma)} \geq \sum_{k=0}^{r} \binom{N}{k}.$$

Then Theorem 3.8 implies that

$$\left|\left\{x \in C_N : \operatorname{dist}(x, B) \leq N/2 - r\right\}\right| \geq \sum_{k=0}^{N/2} \binom{N}{k} = 2^{N-1}.$$

Therefore,

$$\mathbf{P}\left\{x \in C_N : \operatorname{dist}(x, B) \leq \frac{N}{2} - r\right\} \geq \frac{1}{2}.$$

We have that $x = (x_1, \ldots, x_m)$ for some $x_i \in C_n$ and that $\operatorname{dist}(x, B) = \operatorname{dist}(x_1, A) + \ldots + \operatorname{dist}(x_m, A)$ (see Lemma 3.5). Therefore,

$$\mathbf{P}\left\{(x_1, \ldots, x_m) : \frac{1}{m}\sum_{i=1}^{m} \operatorname{dist}(x_i, A) \leq \frac{N}{2m} - \frac{r}{m}\right\} \geq \frac{1}{2}. \tag{1}$$

Now we observe that

$$\frac{N}{2m} - \frac{r}{m} \longrightarrow \frac{n}{2} - n\gamma \quad \text{as} \quad m \longrightarrow +\infty. \tag{2}$$

Furthermore, by the Law of Large Numbers,

$$\frac{1}{m}\sum_{i=1}^{m} \operatorname{dist}(x_i, A) \longrightarrow \Delta(A) \quad \text{in probability} \qquad \text{as} \quad m \longrightarrow +\infty. \tag{3}$$

Hence the assumption that $\Delta(A) > n/2 - n\gamma$ would contradict (1)–(3). Thus we must have $\Delta(A) \leq n/2 - n\gamma$, which implies that $\gamma \leq \rho(A)$. Hence $\alpha = H(\gamma) \leq H(\rho)$ and the upper bound is proven.

Let us prove the lower bound. We observe that for every point $b \in C_N$ and any $N/2 \geq s \geq 0$

$$\left|\{x \in C_N : \operatorname{dist}(x, b) \leq s\}\right| = \sum_{k=0}^{s} \binom{N}{k} \leq 2^{N \cdot H(s/N)}.$$

Therefore,

$$\left|\{x \in C_N : \operatorname{dist}(x, B) \leq s\}\right| \leq |B| 2^{N \cdot H(s/N)} = 2^{N \cdot (H(s/N) + \alpha)}.$$

Hence

$$\mathbf{P}\ \left\{x \in C_N : \operatorname{dist}(x, B) \leq s\right\} \leq 2^{N\cdot(H(s/N)+\alpha-1)}.$$

Therefore,

$$(4) \qquad \mathbf{P}\ \left\{(x_1, \ldots, x_m) : \frac{1}{m}\sum_{i=1}^{m} \operatorname{dist}(x_i, A) \leq s/m\right\} \leq 2^{N\cdot(H(s/N)+\alpha-1)}.$$

If $\Delta(A) = n/2$ then A is a point and the lower bound in Theorem 3.9 is satisfied. Otherwise, let us fix an $\epsilon > 0$ such that $(1+\epsilon)\Delta(A)/n < 1/2$ and let $s = \lceil m(1+\epsilon)\Delta(A)\rceil$. We have

$$(5) \qquad s/m \longrightarrow (1+\epsilon)\Delta(A) \quad \text{and} \quad s/N \longrightarrow (1+\epsilon)\Delta(A)/n \quad \text{as} \quad m \longrightarrow +\infty.$$

Hence the assumption that $H\big((1+\epsilon)\Delta(A)/n\big)+\alpha-1 < 0$ would contradict (3)–(5). Therefore, $H\big((1+\epsilon)\Delta(A)/n\big)+\alpha-1 \geq 0$ for any $\epsilon > 0$ and $H\big(\Delta(A)/n\big)+\alpha-1 \geq 0$. Since $\Delta(A)/n = 0.5-\rho$, the proof follows. □

For applications, the most interesting case is when $n^{-1}\log_2|A|$ is small, that is $\rho \approx 0$.

(3.11) Corollary. *There exist positive constants c_1 and c_2 such that for any non-empty set $A \subset C_n$ and for $\rho = \dfrac{1}{2} - \dfrac{\Delta(A)}{n}$ we have*

$$c_1 \cdot \rho^2 \leq \frac{\ln|A|}{n} \leq c_2 \cdot \rho \ln\frac{1}{\rho}.$$

In particular, for any $c_1 < 2$ and any $c_2 > 1$, the inequality holds in a sufficiently small neighborhood of $\rho = 0$.

Proof. Follows from Theorem 3.9 by (3.7.2). □

(3.12) Discussion. Figure 1 depicts the feasible region for $n^{-1}\log_2|A|$ as de-

scribed by Theorem 3.9.

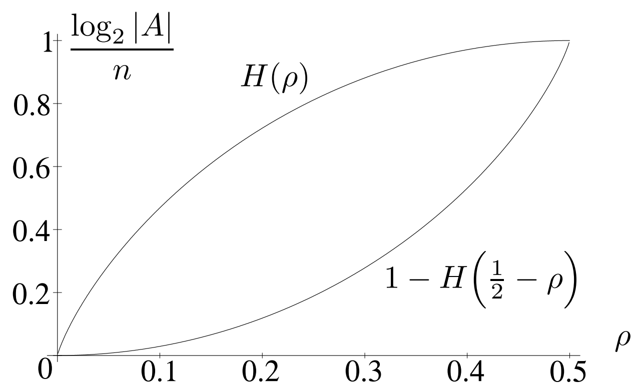

Figure 1

Thus possible values of $n^{-1}\log_2|A|$ with the given value of ρ form a vertical interval between the two curves. As we discussed in Section 3.10, asymptotically both bounds are sharp. Remarkably, the bounds converge at $\rho = 0$ and $\rho = 0.5$. On the other hand, the difference is the greatest when $\rho = 1/4$. Thus, if the average Hamming distance from a point $x \in C_n$ to a set $A \subset C_n$ is $n/4$, the set A can contain as many as $2^{0.811n}$ points and as few as $2^{0.189n}$ points.

Corollary 3.11 (with somewhat weaker constants and stated in different terms) together with the observation that the distance $\operatorname{dist}(x, A)$ for a randomly chosen point $x \in C_n$ allows one to estimate ρ up to an $O(n^{-1/2})$ error constitute the main result of the earlier paper [Barvinok 97a]. Consequently, the main conclusion of [Barvinok 97a] is equivalent to stating that the Hamming distance to A from a random point x in the Boolean cube allows one to decide whether $|A|$ is exponentially large in n. Theorems 3.6 and 3.9 make improvements of two kinds. First, we obtain sharp bounds valid for all $0 \leq \rho \leq 1/2$, and second, by averaging several random distances (see Algorithm 3.3 and Theorem 3.6) we get rid of the $O(n^{-1/2})$ error term. This allows us to obtain meaningful cardinality estimates for really small sets. For example, if $A \subset C_n$ is a set such that $n^{-1}\log_2|A| \sim n^{-\alpha}$, for some $0 < \alpha < 1$, by applying Algorithm 3.3 to approximate $\Delta(A)$ and Theorem 3.9 to interpret the results, the worst lower bound we can get for $n^{-1}\log_2|A|$ is $\sim n^{-2\alpha}\ln^{-2} n$ (this happens when A is a ball in the Hamming metric, but we think it is a "random set", see Section 3.10) and the worst upper bound is $\sim n^{-\alpha/2}\ln n$ (this happens when A is a "random set" but we think that it is a ball). Curiously, we can even distinguish in polynomial time between a set consisting of a single point ($\rho = 0$) and a set having more than one point (one can show that $\rho \geq c/n$ for some $c > 0$

in that case), although apparently we can't distinguish between sets consisting of 2 and 3 points respectively.

As we remarked earlier, in applications the value of $n^{-1}\log_2|A|$ is usually small (cf. Examples 1.2 and 2.5). Therefore, it is of interest to tighten the bounds for such sets. In the next section, we show that this is indeed possible: we demonstrate how to modify the definition of ρ, so that it remains efficiently computable and so that

$$c_3 \cdot \rho^2 \ln\frac{1}{\rho} \leq \frac{\ln|A|}{n} \leq c_4 \cdot \rho \ln\frac{1}{\rho}$$

for some $c_3, c_4 > 0$, which improves the inequality of Corollary 3.11 in the neighborhood of $\rho = 0$.

4. Randomized Hamming Distance

Let us fix a number $0 < p \leq 1$ and let $q = 1 - p$. In this section, we construct a quantity $\Delta(A, p)$, which measures the cardinality of "small" subsets $A \subset C_n$ of the Boolean cube in a somewhat more precise way than the average Hamming Distance $\Delta(A)$ discussed in Section 3. In fact, $\Delta(A, 1) = \Delta(A)$, so $\Delta(A)$ is a particular case of $\Delta(A, p)$.

(4.1) Definitions. Let Λ_n be a copy of the Boolean cube $\{0,1\}^n$. We make Λ_n a probability space by letting

$$\mathbf{P}\ \{l\} = p^{|l|}q^{n-|l|}, \quad \text{where} \quad |l| = \lambda_1 + \ldots + \lambda_n \quad \text{for} \quad l = (\lambda_1, \ldots, \lambda_n).$$

Hence a vector $l = (\lambda_1, \ldots, \lambda_n)$ from Λ_n is interpreted as a realization of n independent random variables λ_i such that $\mathbf{P}\ \{\lambda_i = 1\} = p$ and $\mathbf{P}\ \{\lambda_i = 0\} = q$.

For $x, y \in C_n$ and an $l \in \Lambda_n$, where $x = (\xi_1, \ldots, \xi_n)$, $y = (\eta_1, \ldots, \eta_n)$ and $l = (\lambda_1, \ldots, \lambda_n)$, let

$$d_l(x, y) = \sum_{i:\xi_i \neq \eta_i} \lambda_i.$$

In other words, we count disagreement in the i-th coordinate of x and y if and only if the value of λ_i is 1. Thus if $l = (1, \ldots, 1)$, we have $d_l(x, y) = \operatorname{dist}(x, y)$, the usual Hamming distance.

For $l \in \Lambda_n$ and a set $A \subset C_n$, let

$$d_l(x, A) = \min_{y \in A} d_l(x, y).$$

Finally, let

$$\Delta(A, p) = \sum_{l \in \Lambda_n} \sum_{x \in C_n} d_l(x, A) \frac{p^{|l|}q^{n-|l|}}{2^n}.$$

In other words, $\Delta(A, p)$ is the expected value of $d_l(x, A)$, where $x = (\xi_1, \ldots, \xi_n)$ and $l = (\lambda_1, \ldots, \lambda_n)$ are vectors of independent random variables such that $\mathbf{P}\ \{\lambda_i =$

$1\} = p$, $\mathbf{P}\ \{\lambda_i = 0\} = q$ and $\mathbf{P}\ \{\xi_i = 0\} = \mathbf{P}\ \{\xi_i = 1\} = 1/2$. Obviously, $\Delta(A,p) \leq \Delta(B,p)$ if $B \subset A$.

It follows that for a fixed non-empty $A \subset C_n$, the value $\Delta(A,p)$ is a polynomial in p of degree at most n.

(4.2) Example: Set consisting of a single point. Suppose that the set A consists of a single point. Without loss of generality we assume that $A = \{(0,\ldots,0)\}$. Then for $x = (\xi_1,\ldots,\xi_n)$ and $l = (\lambda_1,\ldots,\lambda_n)$,

$$d_l(x,A) = \sum_{i=1}^{n} \lambda_i \xi_i.$$

Interpreting λ_i and ξ_i, $i = 1,\ldots,n$ as independent random variables such that $\mathbf{P}\ \{\xi_i = 1\} = \mathbf{P}\ \{\xi_i = 0\} = 1/2$ and $\mathbf{P}\ \{\lambda_i = 1\} = p$, $\mathbf{P}\ \{\lambda_i = 0\} = q$, we get

$$\Delta(A,p) = \mathbf{E}\ \sum_{i=1}^{n} \lambda_i \xi_i = \sum_{i=1}^{n} (\mathbf{E}\ \lambda_i)(\mathbf{E}\ \xi_i) = \frac{np}{2}.$$

It follows then that for any non-empty set $A \subset C_n$ we have $\Delta(A,p) \leq np/2$ and that $\Delta(A,p) = np/2$ if and only if A consists of a single point (we agreed that $p > 0$).

As was the case with $\Delta(A)$, the functional $\Delta(A,p)$ can be easily computed by averaging. For a set $A \subset C_n$ defined by its Distance Oracle 2.2 and any $l = (\lambda_1,\ldots,\lambda_n)$ the value of $d_l(x,A)$ is computed by choosing the penalties $d_i(0,1) = d_i(1,0) = 1$ when $\lambda_i = 1$ and $d_i = 0$ when $\lambda_i = 0$.

(4.3) Algorithm for Computing $\Delta(A,p)$

Input: A set $A \subset C_n$ given by its Distance Oracle 2.2, a number $1 \geq p > 0$ and an $\epsilon > 0$.

Output: A number α approximating $\Delta(A,p)$ within error ϵ.

Algorithm: Let $k = \lceil 64n/\epsilon^2 \rceil$. Sample k points $x_1,\ldots,x_k \in C_n$ independently at random from the uniform distribution in C_n and k points $l_1,\ldots,l_k \in \Lambda_n$ independently at random from the distribution in Λ_n. Apply Distance Oracle 2.2 to compute $d_{l_i}(x_i,A)$, $i = 1,\ldots,k$. Compute $\alpha = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k} \mathrm{dist}_{l_i}(x_i,A)$. Output α.

(4.4) Theorem. *With probability at least 0.9, the output α of Algorithm 4.3 satisfies the inequality $|\Delta(A,p) - \alpha| \leq \epsilon$.*

We postpone the proof till Section 6.

We are going to obtain estimates of the cardinality $|A|$ of a set $A \subset C_n$ in terms of the quantity $\Delta(A,p)$. As in Section 3, it is convenient to work with a related quantity

$$\rho = \rho(A,p) = \frac{p}{2} - \frac{\Delta(A,p)}{n}.$$

From Definitions 4.1, for any non-empty $A \subset C_n$, the function $\rho(A,p)$ is a polynomial in p of degree at most n. As follows from Example 4.2, $0 \leq \rho \leq p/2$ for any non-empty set $A \subset C_n$. Our estimate will be useful for "small" sets A where $n^{-1} \ln |A|$ is close to 0.

(4.5) Theorem. *Let $A \subset C_n$ be a non-empty set. Let*

$$\rho = \frac{p}{2} - \frac{\Delta(A,p)}{n}.$$

Then

$$\frac{\rho^2}{p} \leq \frac{\ln |A|}{n}. \tag{4.5.1}$$

Suppose that $\rho \leq 1/4$ and that

$$p \geq \frac{\ln 2 + \ln(1-2\rho)}{\ln(1-2\rho) - \ln(2\rho)}. \tag{4.5.2}$$

Then

$$\frac{\ln |A|}{n} \leq 2\rho \ln \frac{1}{2\rho} + (1-2\rho)\ln \frac{1}{1-2\rho}. \tag{4.5.3}$$

As we remarked earlier, the case interesting for applications is when $|A|$ is small, meaning that $n^{-1} \ln |A| \approx 0$.

(4.6) Corollary. *Let us choose any $c_3 < 1/(\ln 2) \approx 1.44$ and any $c_4 > 2$. Then there exists a $\delta > 0$ such that for any non-empty $A \subset C_n$ with $n^{-1} \ln |A| \leq \delta$ there exists a $0 < p \leq 1$ such that for $\rho = \dfrac{p}{2} - \dfrac{\Delta(A,p)}{n}$ one has*

$$c_3 \cdot \rho^2 \ln \frac{1}{\rho} \leq \frac{\ln |A|}{n} \leq c_4 \cdot \rho \ln \frac{1}{\rho}.$$

Proof. By (4.5.1), $\rho \leq \sqrt{n^{-1} \ln |A|} \leq \sqrt{\delta}$, so $\rho(A,p)$ is small if δ is small, no matter what p is. We observe that for small positive ρ the right hand side of (4.5.2) is of the order $(\ln 2) \ln^{-1}(1/\rho)$ and the right hand side of (4.5.3) is of the order $2\rho \ln(1/\rho)$.

Given $c_3 < (\ln 2)^{-1}$ and $c_4 > 2$, let us choose $1/16 > \delta > 0$ in such a way that the right hand side of (4.5.2) does not exceed $(c_3)^{-1} \ln^{-1}(1/\rho)$ and the right hand side of (4.5.3) does not exceed $c_4 \rho \ln(1/\rho)$ for all $0 < \rho < \sqrt{\delta} < 1/4$.

We recall that $|A| = 1$ if and only if $\rho = 0$, in which case the bounds of Corollary 4.6 are satisfied by default. Given a set $A \subset C_n$, $|A| > 1$, let us choose the smallest $p \geq 0$ that satisfies the inequality (4.5.2). Then $0 < p < 1$ since the right hand side of (4.5.2) is bounded below from 0 as a function of p and n and smaller than 1 for $0 < \rho < 1/4$. Since $\rho(A, p)$ depends continuously on p, we must have equality in (4.5.2) (otherwise, we could have taken a smaller p). Thus $p \leq (c_3)^{-1} \ln^{-1}(1/\rho)$ and the proof follows by (4.5.1)–(4.5.3). $\square$

(4.7) Extremal sets. Let us fix a $0 < p \leq 1$ and an $\epsilon > 0$. Then there exists an $\alpha = \alpha(p, \epsilon) > 0$ with the following property: if $A \subset C_n$ is a set of $\lfloor 2^{\alpha n} \rfloor$ points randomly chosen from the Boolean cube, then with the probability that tends to 1 as n grows to infinity, $n^{-1} \ln |A| < (2 + \epsilon)\rho^2/p$. Hence for any $p > 0$ the bound (4.5.1) is tight up to a constant factor for sufficiently small random sets. The proof is rather technical and therefore omitted.

One can show that as long as p satisfies (4.5.2), the bound (4.5.3) is asymptotically attained on small *faces* of the cube C_n. Let us fix a $\delta > 0$ (to be adjusted later), let $m = \lfloor \delta n \rfloor$ and let $A \subset C_n$ be an m-dimensional face of the Boolean cube:

$$A = \Big\{ (\xi_1, \ldots, \xi_n) : \xi_i = 0 \quad \text{for} \quad i = m + 1, \ldots, n \Big\}.$$

Thus $|A| = 2^m$. Moreover, a computation similar to that of Example 4.2 shows that $\rho(A, p) = pm/2n$. Hence we have

$$\frac{\ln |A|}{n} = \frac{2 \ln 2}{p} \rho(A, p).$$

We observe that $\rho(A, p) \leq \delta/2$. Hence for any small $\epsilon > 0$ one can find $\delta = \delta(\epsilon) > 0$ such that there exists p satisfying (4.5.2) and such that $p < (1 + \epsilon)(\ln 2) \ln^{-1}(1/\rho)$. For such a p, we have

$$\frac{\ln |A|}{n} \geq \frac{2}{1 + \epsilon} \rho \ln \frac{1}{\rho},$$

so the bound (4.5.3) is indeed asymptotically tight for small sets.

Apparently, the sets A having the largest cardinality among all sets with the given value of $\rho(A, p)$ evolve from the balls in the Hamming metric for $p = 1$ (see Section 3.10) to faces at $p \longrightarrow 0$. Since faces are packed somewhat less tightly than balls, we gain in Corollary 4.6 as compared to Corollary 3.11.

The proof of Theorem 4.5 is postponed till Section 6.

Corollary 4.6 implies for small sets A by “tuning up” p we can get an additional logarithmic factor which brings the lower bound for $n^{-1} \ln |A|$ a little closer to the upper bound compared to the bound of Corollary 3.11. Any p which is only

slightly bigger than the bound (4.5.2) will do. For example, if $A \subset C_n$ is a set such that $n^{-1} \ln |A| \sim n^{-\alpha}$ for some $0 < \alpha < 1$, it follows by (4.5.1) that $\rho(A, p) = O(n^{-\alpha/2})$ for any p. Then we can choose some $p = O(\ln^{-1} n)$ that satisfies (4.5.2) (a particular suitable value of p can be found, for example, by dichotomy). Applying Algorithm 4.3 to approximate $\Delta(A, p)$ and Theorem 4.5 to interpret the results, for $n^{-1} \ln |A|$ we would obtain a lower bound of the form $\sim n^{-2\alpha}/\ln n$ at worst and an upper bound of the form $\sim n^{-\alpha/2}\sqrt{\ln n}$ at worst, which is somewhat better than the bounds that could possibly be obtained by using the standard Hamming distance, see Section 3.12.

We are not going to use $\Delta(A, p)$ in what follows, but we find it interesting that some improvement in the cardinality estimate can be achieved by simply ignoring a (random) part of the information contained in the standard Hamming distance.

5. Application: Approximating the Permanent of a 0-1 Matrix

Let $A = (a_{ij})$ be an $n \times n$ matrix. The *permanent* of A is defined by the expression

$$\operatorname{per} A = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)},$$

where S_n is the symmetric group of all substitutions of the set $\{1, \ldots, n\}$. If $a_{ij} \in \{0, 1\}$ for all i and j then $\operatorname{per} A$ counts perfect matchings in a bipartite graph $G_A = (V, E)$, constructed as follows. Let $V = V^+ \cup V^-$ be the set of vertices, where $V^+ = \{1^+, \ldots, n^+\}$ and $V^- = \{1^-, \ldots, n^-\}$, and let $e = (i^+, j^-)$ be an edge of G_A if and only if $a_{ij} = 1$. Then $\operatorname{per} A$ is equal to the number of perfect matchings in G_A, cf. Example 1.2. The problem of computing $\operatorname{per} A$ is # P-hard [Valiant 79] and polynomial time algorithms for computing $\operatorname{per} A$ exactly are known only in few particular cases. For example, if the graph G_A is planar (see [Lovász and Plummer 86]), or more generally, has the genus bounded by some absolute constant [Gallucio and Loebl 99] then $\operatorname{per} A$ can be computed in polynomial time. If the permanent of a 0-1 matrix is small (bounded by a polynomial in the size n of the matrix), it can be computed in polynomial time, see Section 7.3 of [Minc 78] and [Grigoriev and Karpinski 87]. Finally, the permanent of matrices (real or complex) of a small (fixed) rank is computable in polynomial time [Barvinok 96].

Since the exact computation is difficult, the next goal is to find a "very good" approximation algorithm. A fully polynomial time (randomized) approximation scheme is a (probabilistic) algorithm that for any given $\epsilon > 0$ approximates the desired quantity within relative error ϵ in time polynomial in ϵ^{-1}. Probabilistic methods based on rapidly mixing Markov chains resulted in finding such approximation schemes for permanents of dense 0-1 matrices (that is, the matrices with at least $n/2$ 1's in every row and column), random matrices and some special 0-1 matrices (see [Jerrum and Sinclair 89] and [Jerrum and Sinclair 97]). However, for the class of all 0-1 matrices no fully polynomial time randomized approximation

scheme is known (but there is a "mildly exponential" approximation scheme, see [Jerrum and Vazirani 96]).

In [Barvinok 97b], [Barvinok 99] and [Linial *et al.* 20+] a more modest goal was posed and achieved. Given an arbitrary non-negative $n \times n$ matrix A, the polynomial time algorithms [Barvinok 97b] and [Barvinok 99] (randomized) and [Linial *et al.* 20+] (deterministic) produce a number α such that

$$c^n \operatorname{per} A \leq \alpha \leq \operatorname{per} A \tag{5.1}$$

for some absolute constant $c > 0$. Currently the best values of c are $c \approx 0.76$ for the randomized algorithm of [Barvinok 99] and $c \approx 0.37$ for the deterministic algorithm of [Linial *et al.* 20+]. We also note that any polynomial time algorithm achieving a subexponential approximation error can be "upgraded" to a polynomial time approximation scheme, see [Barvinok 99].

Let A be an $n \times n$ matrix of 0's and 1's. If $\operatorname{per} A$ is "big" (for example, if $\operatorname{per} A$ is of the order $n!/2^n$, which is the average value of the permanent for all $n \times n$ 0-1 matrices), the additional factor of c^n in (5.1) should not be considered as a heavy liability. But if $\operatorname{per} A$ is "small" (for example, if $\operatorname{per} A$ is of the order $2^{0.01n}$), the lower bound in (5.1) is useless and the α produced by the algorithms may well be less than 1. The method developed in this section is designed to provide a partial remedy in this situation of a small permanent. Our approach should be considered within the growing family of algorithms that provide a crude yet fast and universally applicable estimates.

Our algorithm for estimating the permanent of a 0-1 matrix A consists of constructing a graph G_A as above, finding an economical embedding of the set $\mathcal{F}$ of perfect matchings in G_A into a Boolean cube (Section 2.4) and estimating the cardinality $|\mathcal{F}|$ using Algorithm 3.3 and Theorem 3.9. We present a summary below.

(5.2) Algorithm for approximating the permanent.

Given an $n \times n$ 0-1 matrix $A = (a_{ij})$, let $G = (V, E)$ be the graph with the set of vertices $V = V^+ \cup V^-$, where $V^+ = \{1^+, \ldots, n^+\}$ and $V^- = \{1^-, \ldots, n^-\}$ and the set of edges $E = \big\{(i^+, j^-) : a_{ij} = 1\big\}$. Let s_i^+ be the degree of i^+ (the i-th row sum of A) and let s_j^- be the degree of j^- (the j-th column sum of A). Let us compute

$$m^+ = \sum_{i=1}^n \lceil \log_2 s_i^+ \rceil \quad \text{and} \quad m^- = \sum_{j=1}^n \lceil \log_2 s_i^- \rceil$$

and let $m = \min(m^+, m^-)$.

Following Section 2.5, we construct an embedding of the set of the perfect matchings in G in $\{0,1\}^m$.

Without loss of generality we assume that $m = m^+$ (otherwise we switch V^+ and V^-, which corresponds to transposing A). Let $m_i = \lceil \log_2 s_i^+ \rceil$, so $m = m_1 + \ldots + m_n$.

To every edge $e = (i^+, j^-)$ of G incident with i^+, let us assign a binary string $\phi_i(e)$ of length m_i so that $\phi_i(e_1) \neq \phi_i(e_2)$ for every pair of distinct edges e_1 and e_2 with the same endvertex i^+.

Given a precision $\epsilon > 0$, let us generate $k = \lceil 48/m\epsilon^2 \rceil$ random binary strings $x_1, \ldots, x_k$, each of length m.

For each $x = x_i$, $i = 1, \ldots, k$, let us do the following procedure:

Consider x as a string of n substrings, $x = y_1 \ldots y_n$, where y_i is a binary string of length m_i. To every edge $e = (i^+, j^-)$ of G assign weight $\gamma_e = \operatorname{dist}\bigl(\phi_i(e), y_i\bigr)$, where dist is the Hamming distance between binary strings. Find the minimum weight $\alpha = \alpha(x)$ of a perfect matching in G using the Assignment Problem algorithm, see Section 11.2 of [Papadimitriou and Steiglitz 98].

Compute the average

$$\alpha = \frac{1}{k} \sum_{i=1}^{k} \alpha(x_i).$$

Compute

$$\beta = \frac{1}{2} - \frac{\alpha}{m}.$$

Output β.

(5.3) Theorem. *Let A be an $n \times n$ 0-1 matrix such that* $\operatorname{per} A > 0$.

Let $s_1^+, \ldots, s_n^+$ be the row and let $s_1^-, \ldots, s_n^-$ be the column sums of A. Let

$$m = \min\left\{ \sum_{i=1}^{n} \lceil \log_2 s_i^+ \rceil, \qquad \sum_{i=1}^{n} \lceil \log_2 s_i^- \rceil \right\}.$$

With probability at least 0.9*, the output β of Algorithm 5.2 satisfies*

$$|\beta - \rho| \leq \epsilon,$$

where $0 \leq \rho \leq 1/2$ is a number such that

$$1 - H\left(\frac{1}{2} - \rho\right) \leq \frac{\log_2 \operatorname{per} A}{m} \leq H(\rho)$$

and $H(x) = x \log_2 \dfrac{1}{x} + (1-x) \log_2 \dfrac{1}{1-x}$ is the entropy function. To find β, Algorithm 5.3 solves $k = \lceil 48/m\epsilon^2 \rceil$ Assignment Problems of size $n \times n$.

Proof. Let $\mathcal{F}$ be the set of perfect matchings in the graph $G = G_A$. The proof follows by the "economical embedding" construction of Section 2.5, Algorithm 3.3, Theorem 3.6 and Theorem 3.9. $\square$

The estimate of Theorem 5.3, however crude, allows us, for example, to decide in polynomial time whether the permanent of a given $n \times n$ 0-1 matrix is subexponential in n. The precise statement is as follows.

(5.4) Corollary. *Let us fix an* $0 < \alpha < 1$ *and let us choose any* $\beta > (1+\alpha)/2$. *Suppose that* A *is* $n \times n$ *0-1 matrix such that* $\operatorname{per} A \leq 2^{n^\alpha}$. *Let us apply Algorithm 5.2 with* $\epsilon = 1/m$. *Then, for all sufficiently large* n, *the estimates of Theorem 5.3 allow us to conclude that* $\operatorname{per} A \leq 2^{n^\beta}$.

Proof. We observe that $m \leq n(\log_2 n + 1)$. By (3.7.2), cf. also Corollary 3.11, we conclude that $\rho = O(n^{\alpha/2} m^{-1/2})$ and the proof follows by Theorem 5.3 and (3.7.2). □

Similarly, one can show that if $\operatorname{per} A \geq 2^{n^\alpha}$ then for any $\beta < 2\alpha - 1$, Algorithm 5.2 with $\epsilon = 1/m$ would allow us to conclude that $\operatorname{per} A \geq 2^{n^\beta}$ for all sufficiently large n. The estimate is, of course, void for $\beta \leq 1/2$, but it is getting better as β approaches 1. For example, if $\operatorname{per} A$ has the order of $2^{n^{0.95}}$, Algorithm 5.2 would allow us to conclude that $\operatorname{per} A$ is greater than $2^{n^{0.89}}$ and is smaller than $2^{n^{0.98}}$.

Corollary 5.4 demonstrates something that none of the exponential error algorithms (cf. (5.1)) can possibly do (neither can any other polynomial time algorithm known to the authors). On the other hand, algorithms of [Barvinok 97b], [Barvinok 99] and [Linial *et al.* 20+] are better than Algorithm 5.2 for matrices with large permanent. Another interesting feature of Algorithm 5.2 is that it clearly favors sparse matrices, as the value of m (the dimension of the cubical embedding, see Example 2.5) for such matrices is smaller. Algorithms from [Barvinok 97b], [Barvinok 99] and [Linial *et al.* 20+] seem to be completely indifferent to sparseness and even show some inclination to like dense matrices better. Thus, in the case of $m = O(n)$ Algorithm 5.2 beats the said algorithms on a wider range of permanents (for example, of the order $2^{0.01n}$). The final remark is about practical implementation of Algorithm 5.2. If $\operatorname{per} A$ is expected to be large enough (say, of the order $2^{\alpha n}$ for some positive α), it suffices to choose $\epsilon = 7m^{-1/2}$, for example. Thus, Algorithm 5.2 boils down to solving one Assignment Problem. The algorithm should be able to handle reasonably sparse matrices with the size n of the order of several hundreds.

Our method applies just as well to counting perfect matchings in non-bipartite graphs, which is a more general problem. We discussed the bipartite case in detail because of its connection with the permanent, a problem with rich history and plenty results available for comparison.

6. Proofs of Theorems 4.4 and 4.5

(6.1) Definition. We recall that C_N is the Boolean cube $\{0,1\}^N$ endowed with the uniform probability measure and that Λ_N is the Boolean cube $\{0,1\}^N$ endowed with the probability measure of Definition 4.1. Let $\Omega_N = C_N \times \Lambda_N$. We consider the product measure on Ω_N, so

$$\mathbf{P}\left\{(x,l)\right\} = p^{|l|} q^{n-|l|} 2^{-N}, \quad \text{where} \quad |l| = \lambda_1 + \ldots + \lambda_N \quad \text{for} \quad l = (\lambda_1, \ldots, \lambda_N).$$

Hence a point $(x,l) \in \Omega_N$ is interpreted as a vector of $2n$ independent random variables $(\xi_1, \ldots, \xi_n; \lambda_1, \ldots, \lambda_n)$, where $\mathbf{P}\left\{\xi_i = 0\right\} = \mathbf{P}\left\{\xi_i = 1\right\} = 1/2$, $\mathbf{P}\left\{\lambda_i = \right.$

$1\} = p$ and $\mathbf{P}\ \{\lambda_i = 0\} = q$. We observe that

$$\Delta(A,p) = \mathbf{E}\ d_l(x,A). \tag{6.1.1}$$

First, we need a version of the concentration inequality (3.4).

(6.2) Lemma. *Let $A \subset C_N$ be a set. Then for every $\delta \geq 0$*

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : |d_l(x,A) - \Delta(A,p)| \geq \delta + 4\sqrt{N}\Big\} \leq 4e^{-\delta^2/N}.$$

Proof. Given an $A \subset C_N$, let $f : \Omega_N \longrightarrow \mathbb{R}$ be defined by $f(x,l) = d_l(x,A)$. Let M be the median of f, that is, a number such that

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : f(x,l) \leq M\Big\} \geq 1/2 \quad \text{and} \quad \mathbf{P}\ \Big\{(x,l) \in \Omega_N : f(x,l) \geq M\Big\} \geq 1/2.$$

Since f is a function with Lipschitz constant 1, it follows by inequality (2.1.3) of [Talagrand 95] that

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : |f(x,l) - M| \geq \delta\Big\} \leq 4e^{-\delta^2/N}$$

for any $\delta \geq 0$.

Since f is integer-valued, we can choose M to be integer. Then

$$\begin{aligned}
\mathbf{E}\ |f(x,l) - M| &= \sum_{k=0}^{+\infty} k\mathbf{P}\ \Big\{(x,l) : |f(x,l) - M| = k\Big\} \\
&= \sum_{k=1}^{+\infty} \mathbf{P}\ \Big\{(x,l) : |f(x,l) - M| \geq k\Big\} \leq 4\sum_{k=1}^{+\infty} e^{-k^2/N} \\
&\leq 4\int_0^{+\infty} e^{-x^2/N} dx = 2\sqrt{\pi N} \leq 4\sqrt{N}.
\end{aligned}$$

Since by (6.1.1) we have $\Delta(A,p) = \mathbf{E}\ f$, we conclude that $|\Delta(A,p) - M| \leq 4\sqrt{N}$. Therefore,

$$\begin{aligned}
\mathbf{P}\ \Big\{(x,l) : |d_l(x,A) - \Delta(A,p)| \geq \delta + 4\sqrt{N}\Big\} &\leq \mathbf{P}\ \Big\{(x,l) : |d_l(x,A) - M| \geq \delta\Big\} \\
&\leq 4\exp\Big\{-\delta^2/N\Big\}.
\end{aligned}$$

$\square$

Next, we need an analogue of the scaling trick (3.5).

(6.3) Lemma. *Let us fix positive integers k and n and let $N = kn$. Let us identify $C_N = (C_n)^k$, $\Lambda_N = (\Lambda_n)^k$ and $\Omega_N = (\Omega_n)^k$. Thus a point $(x, l) \in \Omega_N$ is identified with $x = (x_1, \ldots, x_k; l_1, \ldots, l_k)$, where $x_i \in C_n$ and $l_i \in \Lambda_n$.*

For a subset $A \subset C_n$, let $B = A^k \subset C_N$. Then

$$d_l(x, B) = \sum_{i=1}^{k} d_{l_i}(x_i, A) \quad \textit{and} \quad \Delta(B, p) = k\Delta(A).$$

Proof. Clearly,

$$d_l(x, y) = \sum_{i=1}^{k} d_{l_i}(x_i, y_i) \quad \text{for all} \quad x, y \in C_N$$

and the first identity follows. Now, by (6.1.1)

$$\Delta(B, p) = \mathbf{E}\ d_l(x, B) = \sum_{i=1}^{k} \mathbf{E}\ d_{l_i}(x_i, A) = k\Delta(A, p).$$

□

Now we are ready to prove Theorem 4.4.

Proof of Theorem 4.4. Let $N = nk$ and let us identify $C_N = (C_n)^k$, $\Lambda_N = (\Lambda_n)^k$ and $\Omega_N = (\Omega_n)^k$. Let $B = A^k \subset C_N$ as in Lemma 6.3. Applying Lemma 6.2, we get

$$\mathbf{P}\ \left\{(x, l) \in \Omega_N : |d_l(x, B) - \Delta(B, p)| \geq \delta + 4\sqrt{N}\right\} \leq 4e^{-\delta^2/N}$$

for any $\delta \geq 0$. Using Lemma 6.3, we conclude:

$$\mathbf{P}\ \left\{(x, l) \in \Omega_N : \Big|\frac{1}{k}\sum_{i=1}^{k} d_{l_i}(x_i, A) - \Delta(A, p)\Big| \geq \delta/k + 4\sqrt{n/k}\right\} \leq 4e^{-\delta^2/N}.$$

Let us choose $\delta = \epsilon k/2$. Hence

$$\mathbf{P}\ \left\{(x, l) \in \Omega_N : \Big|\frac{1}{k}\sum_{i=1}^{k} d_{l_i}(x_i, A) - \Delta(A, p)\Big| \geq \epsilon/2 + 4\sqrt{n/k}\right\} \leq 4e^{-\epsilon^2 k/n}.$$

Since $k \geq 64n/\epsilon^2$, the proof follows. □

Next, we need a (crude) version of inequality (3.7.1).

(6.4) Lemma. *Let $\epsilon \geq 0$, let $r(\epsilon) = pN(1-\epsilon)/2$. Let $y \in C_N$ be a point. Then*

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,y) \leq r(\epsilon)\Big\} \leq e^{-\epsilon^2 pN/4}.$$

Proof. Without loss of generality we may assume that $y = 0$. Then

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,0) \leq r(\epsilon)\Big\} = \mathbf{P}\ \Big\{(x,l) \in \Omega_N : \sum_{i=1}^{N} \xi_i \lambda_i \leq r(\epsilon)\Big\},$$

where $x = (\xi_1, \dots, \xi_N)$ and $l = (\lambda_1, \dots, \lambda_N)$. Let $\zeta_i = \xi_i \lambda_i$. Then ζ_i, $i = 1, \dots, N$ are independent random variables such that $\mathbf{P}\ \{\zeta_i = 1\} = p/2$ and $\mathbf{P}\ \{\zeta_i = 0\} = 1 - p/2$. Hence

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,y) \leq r(\epsilon)\Big\} = \mathbf{P}\ \Big\{\zeta_1 + \ldots + \zeta_N \leq r(\epsilon)\Big\} \leq e^{-\epsilon^2 pN/4}$$

by a corollary of Chernoff's inequality (see [McDiarmid 89]). □

Now we are ready to prove the first part of Theorem 4.5.

Proof of inequality (4.5.1). Let us choose a positive integer m, let $N = mn$, let $C_N = (C_n)^m$, and let $\Lambda_N = (\Lambda_n)^m$. Let $B = A^m \subset C_N$ as in Lemma 6.3.

Let us choose an $\alpha > 0$. Applying Lemma 6.4, we obtain

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,B) \leq pN(1-\sqrt{\alpha})/2\Big\} \leq |B| e^{-\alpha pN/4} = (|A| e^{-\alpha pn/4})^m.$$

Therefore, by Lemma 6.3

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : \frac{1}{m} \sum_{i=1}^{m} d_{l_i}(x_i, A) \leq pn(1-\sqrt{\alpha})/2\Big\} \leq (|A| e^{-\alpha pn/4})^m.$$

The right hand side of the inequality tends to 0 provided $\alpha > 4 \ln |A|/pn$. Since by the Law of Large Numbers

$$\frac{1}{m} \sum_{i=1}^{m} d_{l_i}(x_i, A) \longrightarrow \Delta(A,p) \quad \text{in probability} \quad \text{as} \quad m \longrightarrow +\infty,$$

we must have

$$\Delta(A,p) \geq pn(1-\sqrt{\alpha})/2 \quad \text{for any} \quad \alpha > 4 \ln |A|/pn.$$

Hence

$$\Delta(A,p) \geq pn(1-\sqrt{\alpha})/2 \quad \text{for} \quad \alpha = 4 \ln |A|/pn,$$

which is equivalent to (4.5.1). □

In Section 3, we used the sharp isoperimetric inequality (Theorem 3.8) for the Hamming distance in C_n to get a sharp upper bound for $n^{-1} \log_2 |A|$. Unfortunately, we don't know of a similar result for the randomized Hamming distance. To prove (4.5.2)–(4.5.3), we proceed by induction on n in a way resembling that of [Talagrand 95] (see also Remark 6.9).

We start with a simple technical result.

(6.5) Lemma. *For any $0 \le \epsilon \le 1$, any $\gamma \ge 0$ and any $0 < p \le 1$ and $q = 1 - p$ we have*

$$\min\Big\{\frac{p\gamma}{2} + \ln\frac{1}{1+\epsilon}, \quad p\ln\frac{1}{1-\epsilon} + q\ln\frac{1}{1+\epsilon}\Big\} \le \max\Big\{0, \quad \ln(1+e^{\gamma/2}) - \frac{q\gamma}{2} - \ln 2\Big\}.$$

Proof. Fixing p, q and γ, let

$$f(\epsilon) = \frac{p\gamma}{2} + \ln\frac{1}{1+\epsilon} \quad \text{and} \quad g(\epsilon) = p\ln\frac{1}{1-\epsilon} + q\ln\frac{1}{1+\epsilon}.$$

Then $f(0) \ge 0$ and $f(\epsilon)$ is decreasing whereas $g(\epsilon)$ behaves as follows: $g(0) = 0$ and if $p \ge q$ then $g(\epsilon)$ is increasing and if $p < q$ then $g(\epsilon)$ is decreasing for $0 < \epsilon < q - p$ and increasing for $q - p < \epsilon < 1$. Furthermore, $f(\epsilon_0) = g(\epsilon_0)$ at the single point $\epsilon_0 = (e^{\gamma/2} - 1)/(1 + e^{\gamma/2})$, where $f(\epsilon_0) = g(\epsilon_0) = \ln(1 + e^{\gamma/2}) - q\gamma/2 - \ln 2$. The proof now follows. □

(6.6) Definition. Let μ_n (or simply μ) denote the uniform probability measure in C_n. Hence $\mu(A) = |A|/2^n$.

The induction is based on the following lemma.

(6.7) Lemma. *Let $A \subset C_{n+1}$ be a set. Let*

$$A_0 = \big\{x \in C_n : (x, 0) \in A\big\} \quad \textit{and} \quad A_1 = \big\{x \in C_n : (x, 1) \in A\big\}.$$

For $l \in \Lambda_n$ let $(l, 0) \in \Lambda_{n+1}$ denote l appended by $\lambda_{n+1} = 0$ and let $(l, 1) \in \Lambda_{n+1}$ denote l appended by $\lambda_{n+1} = 1$. Let

$$\Delta_0(A, p) = \mathbf{E}\ d_{(l,0)}(x, A) \quad \textit{and} \quad \Delta_1(A, p) = \mathbf{E}\ d_{(l,1)}(x, A),$$

where the expectation is taken with respect to a random $(x, l) \in C_{n+1} \times \Lambda_n$. Then

$$\frac{\mu_n(A_0) + \mu_n(A_1)}{2} = \mu_{n+1}(A); \tag{6.7.1}$$

$$\Delta(A, p) = q\Delta_0(A, p) + p\Delta_1(A, p); \tag{6.7.2}$$

$$\Delta_0(A, p) \le \Delta(A_i, p) \quad \textit{for} \quad i = 0, 1; \tag{6.7.3}$$

$$\Delta_1(A, p) \le \Delta(A_i, p) + \frac{1}{2} \quad \textit{for} \quad i = 0, 1; \tag{6.7.4}$$

$$\Delta_1(A, p) \le \frac{\Delta(A_0, p) + \Delta(A_1, p)}{2}. \tag{6.7.5}$$

Proof. Clearly, $|A_0| + |A_1| = |A|$, so (6.7.1) follows. Identity (6.7.2) is immediate from Definitions 4.1. We observe that for any $x, y \in C_n$,

$$d_{(l,0)}((x,j),(y,i)) = d_l(x,y), \quad \text{where} \quad i,j \in \{0,1\}.$$

Hence

$$d_{(l,0)}((x,j),A) \le d_l(x,A_i), \quad i,j = 0,1$$

and (6.7.3) follows by averaging.

Next, we observe that

$$d_{(l,1)}((x,i),(y,j)) = \begin{cases} d_l(x,y) & \text{if } i = j \\ d_l(x,y) + 1 & \text{if } i \ne j. \end{cases}$$

Therefore,

$$d_{(l,1)}((x,1),A) = \min\bigl\{d_l(x,A_1), d_l(x,A_0) + 1\bigr\}$$

and

$$d_{(l,1)}((x,0),A) = \min\bigl\{d_l(x,A_0), d_l(x,A_1) + 1\bigr\}.$$

Averaging over $(x,l) \in C_{n+1} \times \Lambda_n$, we get

$$\begin{aligned} \Delta_1(A,p) = \mathbf{E}\ d_{(l,1)}(x,A) &= \frac{\mathbf{E}\ d_{(l,1)}\bigl((x,1),A\bigr) + \mathbf{E}\ d_{(l,1)}\bigl((x,0),A\bigr)}{2} \\ &\le \frac{\mathbf{E}\ d_l(x,A_1) + \mathbf{E}\ d_l(x,A_1) + 1}{2} = \Delta(A_1,p) + \frac{1}{2}. \end{aligned}$$

Similarly,

$$\begin{aligned} \Delta_1(A,p) = \mathbf{E}\ d_{(l,1)}(x,A) &= \frac{\mathbf{E}\ d_{(l,1)}\bigl((x,1),A\bigr) + \mathbf{E}\ d_{(l,1)}\bigl((x,0),A\bigr)}{2} \\ &\le \frac{\mathbf{E}\ d_l(x,A_0) + 1 + \mathbf{E}\ d_l(x,A_0)}{2} = \Delta(A_0,p) + \frac{1}{2}, \end{aligned}$$

which completes the proof of (6.7.4). Finally,

$$\begin{aligned} \Delta_1(A,p) = \mathbf{E}\ d_{(l,1)}(x,A) &= \frac{\mathbf{E}\ d_{(l,1)}\bigl((x,1),A\bigr) + \mathbf{E}\ d_{(l,1)}\bigl((x,0),A\bigr)}{2} \\ &\le \frac{\mathbf{E}\ d_l(x,A_1) + \mathbf{E}\ d_l(x,A_0)}{2} = \frac{\Delta(A_1,p) + \Delta(A_0,p)}{2} \end{aligned}$$

and (6.7.5) is proved. $\square$

Now we use induction to get a preliminary bound.

(6.8) Lemma. *Suppose that for some $\gamma \geq 0$, $0 < p \leq 1$ and $q = 1 - p$,*

$$\ln\bigl(1 + e^{\gamma/2}\bigr) - \frac{q\gamma}{2} - \ln 2 \geq 0.$$

Then for any non-empty set $A \subset C_n$ we have

$$\gamma\Delta(A, p) + \ln \mu(A) \leq n\Bigl(\ln\bigl(1 + e^{\gamma/2}\bigr) - \frac{q\gamma}{2} - \ln 2\Bigr).$$

Proof. We proceed by induction on n. If $n = 1$ then the two cases are possible:

A consists of a single point, $\mu(A) = 1/2$ and $\Delta(A, p) = p/2$ (see Example 4.2);

$A = \{0, 1\}$, $\mu(A) = 1$ and $\Delta(A, p) = 0$.

In both cases the inequality holds.

Suppose that the inequality holds for non-empty subsets of C_n. Let us prove that it holds for non-empty $A \subset C_{n+1}$. Let us define $A_0, A_1 \subset C_n$ as in Lemma 6.7. From (6.7.1) it follows that either

$$\mu_n(A_0) = (1 - \epsilon)\mu_{n+1}(A) \quad \text{and} \quad \mu_n(A_1) = (1 + \epsilon)\mu_{n+1}(A)$$

or

$$\mu_n(A_1) = (1 - \epsilon)\mu_{n+1}(A) \quad \text{and} \quad \mu_n(A_0) = (1 + \epsilon)\mu_{n+1}(A)$$

for some $0 \leq \epsilon \leq 1$.

Let B be the one of the sets A_0, A_1 that has a bigger measure μ_n (either of the two if $\mu_n(A_0) = \mu_n(A_1)$) and let D be the one of the sets A_0, A_1 that has a bigger value of $\Delta(\cdot, p)$ (either of the two if $\Delta(A_0, p) = \Delta(A_1, p)$). Then

$$\mu_n(B) \geq (1 + \epsilon)\mu_{n+1}(A) \quad \text{and} \quad \mu_n(D) \geq (1 - \epsilon)\mu_{n+1}(A).$$

Furthermore, by (6.7.3)

$$\Delta_0(A, p) \leq \Delta(B, p) \quad \text{and} \quad \Delta_0(A, p) \leq \Delta(D, p)$$

whereas by (6.7.3) and (6.7.5)

$$\Delta_1(A, p) \leq \Delta(B, p) + \frac{1}{2} \quad \text{and} \quad \Delta_1(A, p) \leq \Delta(D, p).$$

Hence we get

$$\gamma\Delta_0(A, p) + \ln \mu_{n+1}(A) \leq \gamma\Delta(B, p) + \ln \mu_n(B) + \ln \frac{1}{1 + \epsilon}$$

and

$$\gamma\Delta_1(A,p)+\ln\mu_{n+1}(A)\le$$
$$\min\Big\{\gamma\Delta(B,p)+\ln\mu_n(B)+\ln\frac{1}{1+\epsilon}+\frac{\gamma}{2},\quad \gamma\Delta(D,p)+\ln\mu_n(D)+\ln\frac{1}{1-\epsilon}\Big\}.$$

Clearly, B is non-empty. Assume first, that D is non-empty as well. Applying the induction hypothesis to B and D, we conclude that

$$\gamma\Delta_0(A,p)+\ln\mu_{n+1}(A)\le n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\ln\frac{1}{1+\epsilon}$$

and

$$\gamma\Delta_1(A,p)+\ln\mu_{n+1}(A)\le n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\min\Big\{\ln\frac{1}{1+\epsilon}+\frac{\gamma}{2},\quad \ln\frac{1}{1-\epsilon}\Big\}.$$

Adding the first inequality multiplied by q and the second inequality multiplied by p and using (6.7.2), we get

$$\gamma\Delta(A,p)+\ln\mu_{n+1}(A)\le$$
$$n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\min\Big\{\frac{p\gamma}{2}+\ln\frac{1}{1+\epsilon},\quad p\ln\frac{1}{1-\epsilon}+q\ln\frac{1}{1+\epsilon}\Big\}.$$

The desired inequality follows by Lemma 6.4.

If D is empty then $\mu_n(B)=2\mu_{n+1}(A)$ and we obtain

$$\gamma\Delta_0(A,p)+\ln\mu_{n+1}(A)\le\gamma\Delta(B,p)+\ln\mu_n(B)-\ln 2$$

and

$$\gamma\Delta_1(A,p)+\ln\mu_{n+1}(A)\le\gamma\Delta(B,p)+\ln\mu_n(B)-\ln 2+\frac{\gamma}{2}$$

Adding the first inequality multiplied by q to the second inequality multiplied by p and using (6.7.2) and the induction hypothesis, we get:

$$\begin{aligned}\gamma\Delta(A,p)+\ln\mu_{n+1}(A)&\le\gamma\Delta(B,p)+\ln\mu_n(B)-\ln 2+\frac{p\gamma}{2}\\&\le n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\Big(\frac{\gamma}{2}-\frac{q\gamma}{2}-\ln 2\Big)\\&\le(n+1)\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big),\end{aligned}$$

which completes the proof. $\square$

Now we are ready to complete the proof of Theorem 4.5.

Proof of (4.5.2)–(4.5.3). By Lemma 6.8,

$$\frac{\ln|A|}{n}=\frac{\ln\mu_n(A)}{n}+\ln 2\le\ln(1+e^{\gamma/2})-\frac{q\gamma}{2}-\frac{\gamma\Delta(A,p)}{n}=\ln(1+e^{\gamma/2})-\frac{\gamma}{2}+\gamma\rho$$

provided

$$\ln\bigl(1+e^{\gamma/2}\bigr)-\frac{q\gamma}{2}-\ln 2\geq 0.$$

We optimize the inequality on $\gamma \geq 0$. Let

$$\gamma = 2\ln\Bigl(\frac{1}{2\rho}-1\Bigr).$$

Since we assumed that $\rho \leq 1/4$, we have $\gamma \geq 0$. Furthermore,

$$\begin{aligned}\ln\bigl(1+e^{\gamma/2}\bigr)-\frac{q\gamma}{2}-\ln 2 &= \ln\frac{1}{2\rho}-q\ln\Bigl(\frac{1}{2\rho}-1\Bigr)-\ln 2\\ &= -\ln(1-2\rho)+p\bigl(\ln(1-2\rho)-\ln(2\rho)\bigr)-\ln 2\geq 0,\end{aligned}$$

because of (4.5.2). Therefore,

$$\frac{\ln|A|}{n}\leq \ln\frac{1}{2\rho}-\ln\frac{1-2\rho}{2\rho}+2\rho\ln\frac{1-2\rho}{2\rho}=2\rho\ln\frac{1}{2\rho}+(1-2\rho)\ln\frac{1}{1-2\rho}$$

and (4.5.3) follows. □

(6.9) Remark. Our proof of (4.5.2)–(4.5.3) can be considered as an "additive" version of Talagrand's method [Talagrand 95]. Indeed, Talagrand's approach very roughly can be can be stated as follows. Let Ω be a space with the distance function d and probability measure μ. To prove an isoperimetric inequality for $A \subset \Omega$, we first find a uniform bound for the expression $\mu^{\alpha}(A)\cdot \mathbf{E}\ \exp\{\tau d(x,A)\}$ and then adjust parameters $\alpha > 0$ and $\tau > 0$. This way tight inequalities are obtained in [Talagrand 95] for sets A of large measure, most often with $\mu(A) \geq 1/2$. We are mostly interested in sets of a small measure. One can check that for "small sets" A the inequalities of [Talagrand 95] are very far from sharp, which is, of course, should not be perceived as a "fault" of the method, since the method was designed for totally different problems. We find a uniform bound for the expression $\ln\mu(A)+\alpha\mathbf{E}\ d(x,A)$, which looks like Talagrand' functional with "exp" removed. Our method seems to produce reasonably good bounds for small sets A but it fails miserably for large A, with $\mu(A)=1/2$, say. As should have been expected, the case of "middle-sized" sets is the most complicated.

7. Concluding Remarks

Connections to Monte-Carlo methods. The main idea of our approach can be described as follows: given a (finite) ambient space Ω and a set $A \subset \Omega$, we estimate the cardinality $|A|$ by choosing a certain distance function d in Ω and estimating the average distance

$$\Delta(A)=\frac{1}{|\Omega|}\sum_{x\in\Omega}d(x,A),\quad \text{where}\quad d(x,A)=\min_{y\in A}d(x,y)$$

from $x \in \Omega$ to A. We get the classical Monte-Carlo method if the distance function d is the simplest possible:

$$d(x,y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } \ x = y. \end{cases}$$

In this case, $\Delta(A) = |A|/|\Omega|$, so there is a direct relation between $\Delta(A)$ and $|A|$. It is well understood that the main difficulty with the Monte-Carlo method is that if $|A|$ is very small compared to $|\Omega|$, it is hard to get an estimate for the cardinality of A different from 0. In other words, if $|A|$ is "exponentially small" compared to $|\Omega|$, to get a non-trivial bound for $|A|$, we have to compute $\Delta(A)$ with exponentially high precision. In this paper, we showed that in many interesting cases one can choose a different distance function d, so that the distance $d(x, A)$ from a point $x \in \Omega$ to A is efficiently computable and to get a meaningful estimate of $|A|$ even for exponentially small sets A, one need to compute $\Delta(A)$ with a polynomial precision.

Hence our approach can be considered as a natural extension of the Monte-Carlo method. In this context, economical embedding 2.4 can be considered as an analogue of the "importance sampling", whose objective is to replace a large ambient space Ω by a smaller space containing A.

Embedding in different metric spaces. Given a combinatorially defined family $\mathcal{F} \subset 2^X$, we constructed its embedding into the Boolean cube $\{0,1\}^n$ and investigated what happens in the cube is endowed either with the standard Hamming distance (Section 3) or with its randomized version (Section 4). In many cases, there are different ways of metrization of $\mathcal{F}$. One example is provided by the set $\mathcal{F}$ of perfect matchings in a given bipartite graph studied in the paper.

Let $G = (V^+ \cup V^-, E)$ be a bipartite graph with $V^+ = \{1^+, \ldots, n^+\}$, $V^- = \{1^-, \ldots, n^-\}$ (cf. Example 1.2 and Section 5). For every vertex $i^+ \in V_+$, let

$$\Omega_i = \left\{ e = (i^+, j^-) : e \in E \right\}$$

be the set of edges of G coming out of i^+. Let

$$\Omega = \Omega_1 \times \ldots \times \Omega_n.$$

Every perfect matching in G can be identified with a point in Ω, so the set $\mathcal{F}$ of all perfect matchings in G is identified with a subset $F \subset \Omega$.

Let d_i be a distance function on Ω_i, $i = 1, \ldots, n$. Let us define the distance function d on Ω by

$$d(x,y) = \sum_{i=1}^{n} d_i(x_i, y_i), \quad \text{where} \quad x = (x_1, \ldots, x_n) \quad \text{and} \quad y = (y_1, \ldots, y_n).$$

It is easy to check that for any $x \in \Omega$, $x = (x_1, \ldots, x_n)$ the distance $d(x, F)$ is the minimum weight of a perfect matching in G with weighting $\gamma(e) = d_i(e, x_i)$ for

$e = (i^+, j^-)$. Hence for any x, the value of $d(x, F)$ can be found in $O(n^3)$ time. How should we choose d_i to get the best possible estimates for the number $|F|$ of perfect matchings in G?

The authors looked into some of the most obvious candidates, when d_i is a graph metric on Ω_i for a complete graph and for a path (circle). Interestingly, choosing Ω_i isometric to a subset of a power of the complete graph with $m = \Theta(\ln n)$ vertices leads to an improvement by logarithmic factor similar to that of the randomized Hamming distance (Section 4). The choice of d_i used in this paper comes from identifying Ω_i with a subset of the Boolean cube $\{0,1\}^{m_i}$ for $m_i = \lceil \log_2 |\Omega_i| \rceil$, see Section 5. Perhaps one should use a whole family of distance functions d_i and combine the resulting estimates. General isoperimetric inequality of [Alon *et al.* 98] may be very useful for that.

Weighted counting. Let $\mathcal{F} \subset 2^X$ be a family of subsets of the ground set $X = \{1, \dots, n\}$ and let $\mu(i) = p_i/q_i > 0$ be a rational weight of $i \in X$, where $p_i, q_i \in \mathbb{N}$. Let us define

$$\mu(Y) = \prod_{i \in Y} \mu(i) \quad \text{for} \quad Y \in \mathcal{F} \qquad \text{and} \qquad \mu(\mathcal{F}) = \sum_{Y \in \mathcal{F}} \mu(Y).$$

We may be interested to estimate $\mu(\mathcal{F})$. There are several ways to extend our methods to problems of this type, here we sketch one. For every $i \in X$, let $m_i = \lceil \log_2(p_i + q_i) \rceil$. Let us choose subsets $A_i \subset C_{m_i}$ and $B_i \subset C_{m_i}$ such that $|A_i| = p_i$, $|B_i| = q_i$ and $A_i \cap B_i = \emptyset$. Let $m = m_1 + \ldots + m_n$ and let us identify

$$C_m = C_{m_1} \times \ldots \times C_{m_n}.$$

For $Y \subset \mathcal{F}$ let $Z_Y \subset C_m$ be the direct product of n factors, the i-th factor being A_i if $i \in Y$ and B_i if $i \notin Y$. Finally, let $F \subset C_m$ be the union of all Z_Y for $Y \in \mathcal{F}$. We see that $\mu(\mathcal{F}) = (q_1 \cdots q_n)^{-1}|F|$. Moreover, one can define subsets A_i and B_i in such a way that Optimization Oracle 1.1 for $\mathcal{F}$ gives rise to Distance Oracle 2.2 for F. This construction corresponds to the straightforward embedding (2.3). In some cases, there is a way to come up with an economical embedding in the spirit of (2.4).

Acknowledgment

The authors are grateful to M. Gromov and B. Sudakov for many helpful discussions.

References

[Alon *et al.* 98] N. Alon, R. Boppana and J. Spencer, An asymptotic isoperimetric inequality, *Geom. Funct. Anal.* **8** (1998), 411–436.

[Barvinok 96] A. Barvinok, Two algorithmic results for the traveling salesman

problem, *Math. Oper. Res.* **21** (1996), 65–84.

[Barvinok 97a] A. Barvinok, Approximate counting via random optimization, *Random Structures & Algorithms*, **11**(1997), no. 2, 187–198.

[Barvinok 97b] A. Barvinok, Computing mixed discriminants, mixed volumes, and permanents, *Discrete Comput. Geom.* **18** (1997), no. 2, 205–237.

[Barvinok 99] A. Barvinok, Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor, *Random Structures & Algorithms*, **14**(1999), no. 1, 29–61.

[Gallucio and Loebl 99] A. Galluccio and M. Loebl, On the theory of Pfaffian orientations. I. Perfect matchings and permanents, *Electron. J. Combin.* **6** (1999), no. 1, Research Paper 6, 18 pp. (electronic).

[Grigoriev and Karpinski 87], D. Yu Grigoriev and M. Karpinski, The matching problem for bipartite graphs with polynomially bounded permanents is in NC, *Proc. Twenty-Eight Annual IEEE Symp. Foundations of Computer Science*, IEEE, Computer Society Press, Washington, DC, 1987, 162–172.

[Jerrum 95] M. Jerrum, The computational complexity of counting, *Proceedings of the International Congress of Mathematicians*, Vol. 1,2 (Zürich, 1994), 1407–1416, Birkhäuser, Basel, 1995.

[Jerrum and Sinclair 89] M. Jerrum and A. Sinclair, Approximating the permanent, *SIAM J. Comput.*, **18** (1989), no. 6, 1149–1178.

[Jerrum and Sinclair 97] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo Method: an approach to approximate counting and integration, in: *Approximation Algorithms for NP-hard Problems*, ed. D.S. Hochbaum, PWS, Boston, 1997, 483–520.

[Jerrum *et al.* 86] M. Jerrum, L.G. Valiant and V.V. Vazirani, Random generation of combinatorial structures from a uniform distribution, *Theoret. Comput. Sci.*, **43** (1986), no. 2-3, 169–188.

[Jerrum and Vazirani 96] M. Jerrum and U. Vazirani, A mildly exponential approximation algorithm for the permanent, *Algorithmica*, **16**(1996), no. 4-5, 392–401.

[Leader 91] I. Leader, Discrete isoperimetric inequalities, in: *Probabilistic Combinatorics and its Applications (San Francisco, CA, 1991)*, 57–80, Proc. Sympos. Appl. Math., 44, Amer. Math. Soc., Providence, RI, 1991.

[Linial *et al.* 20+] N. Linial, A. Samorodnitsky and A. Wigderson, A deterministic strongly polynomial algorithm for matrix scaling and approximate permanents, *Combinatorica*, to appear.

[van Lint 99] J.H. van Lint, *Introduction to Coding Theory*, Third edition. Graduate Texts in Mathematics, **86**, Springer-Verlag, Berlin, 1999.

[Lovász and Plummer 86] L. Lovász and M.D. Plummer, *Matching Theory*, North-Holland Mathematics Studies, **121**, Annals of Discrete Mathematics, **29**,

North-Holland Publishing Co., Amsterdam-New York; Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences), Budapest, 1986.

[McDiarmid 89] C. McDiarmid, On the method of bounded differences, in: *Surveys in Combinatorics*, 1989 (Norwich, 1989), 148–188, London Math. Soc. Lecture Note Ser., **141**, Cambridge Univ. Press, Cambridge, 1989

[Milman and Schechtman 86] V.D. Milman and G. Schechtman, *Asymptotic Theory of Finite-Dimensional Normed Spaces. With an Appendix by M. Gromov*, Lecture Notes in Mathematics, **1200**, Springer-Verlag, Berlin-New York, 1986.

[Minc 78] H. Minc, *Permanents*, Encyclopedia of Mathematics and its Applications, **6**, Addison-Wesley, Reading, Mass., 1978.

[Papadimitriou and Steiglitz 98] C.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Dover, NY, 1998.

[Talagrand 95] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci. Publ. Math.* No. 81, (1995), 73–205.

[Valiant 79] L.G. Valiant, The complexity of computing the permanent, *Theoret. Comput. Sci.*, **8** (1979), no. 2, 189–201.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109
E-mail address: `barvinok@math.lsa.umich.edu`

Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540
E-mail address: `asamor@ias.edu`